



CLASIFICACIÓN DE NIVELES DE ESTRÉS UTILIZANDO SEÑALES FISIOLÓGICAS

Autor: Sergio Alcázar Rentero

Director: Antonio Artés Rodríguez

Codirector: Francisco Hernando Gallego

Dto. de Teoría de la Señal y Comunicaciones

UNIVERSIDAD CARLOS III DE MADRID

Índice

1	Introducción	6
1.1	Motivación	6
1.2	Objetivos	6
1.3	Entorno socio-económico	7
1.4	Marco regulatorio	7
1.5	Estructura del proyecto	7
2	Estado del arte	8
2.1	Actividad cardíaca	10
2.2	Actividad electrodérmica	12
2.3	Dilatación de las pupilas	13
3	Planteamiento del problema	14
3.1	Diseño del experimento	14
3.1.1	Eriksen Flanker Task	15
3.1.2	Math Processing Task	15
3.1.3	Clock Test	16
3.1.4	Choice Response Time Task	17
3.2	Adquisición de datos	18
3.2.1	Electrocardiogram	18
3.2.2	Galvanic Skin Response	19
4	Procesado de señales	21
4.1	Procesado de los datos	21
4.1.1	ECG	22
4.1.2	GSR	23
4.2	Extracción de características	25
4.2.1	Actividad cardíaca	25
4.2.2	Actividad Electrodérmica	26
5	Resultados	27
5.1	Clasificación del estado de ansiedad y rendimiento	27
5.2	Niveles de estrés	28
5.3	Regresión logística	29
5.4	K-Nearest Neighbors	31
6	Conclusiones y trabajo futuro	33
6.1	Conclusiones	33
6.2	Trabajo futuro	33

7	Presupuesto y planificación del trabajo	35
7.1	Planificación	35
7.2	Presupuesto	35
A	Lista de Acrónimos	37
B	Summary: Classification of stress levels applying physiological signals	38
B.1	Introduction	38
B.2	Experiment design	38
B.2.1	Experiment	39
B.2.2	Data acquisition	41
B.3	Data processing	42
B.3.1	Preprocessing	42
B.3.2	Feature extraction	43
B.4	Results	44
B.5	Conclusions and future work	46

Índice de figuras

1	Curva de Yerkes-Dodson	9
2	Ejemplo de señal BVP	11
3	Ejemplo de señal ECG	11
4	Ejemplo de señal GSR	12
5	Estructura del experimento	14
6	Escenarios del test Eriksen Flanker Task	15
7	Escenarios del test Math Processing Task	16
8	Escenario del Clock Test	16
9	Escenario del Clock Test con solapamiento	17
10	Fases de estímulo y respuesta del Choice Response Time Task	17
11	Colocación del sensor <i>SHIMMER3 ECG</i>	19
12	Colocación del sensor <i>Q sensor</i>	20
13	Respuestas correctas e incorrectas a lo largo de una prueba	21
14	Media de las respuestas en una ventana de cinco muestras.	22
15	Media de las respuestas suavizadas con un FPB	22
16	Respuestas incorrectas a lo largo de una prueba	22
17	Ejemplo de señal de actividad electrodérmica	23
18	Procesado de señal GSR	24
19	HR y HR medio	25
20	HR y HRV	26
21	HR y pendiente del HR	26
22	Pendiente del GSR	26
23	Separación entre niveles de estrés en un test	27
24	Niveles de estrés a lo largo del experimento	28
25	Acierto en función del parámetro K	31
1	Yerkes-Dodson curve	39
2	Experiment structure	39
3	Eriksen Flanker Task Scenarios	40
4	Math Processing Task Scenarios	40
5	Clock Test scenario	41
6	Choice Response Time test phases	41
7	Sensor placement	42
8	Stress levels	44

Índice de tablas

1	Datos extraídos para cada participante	18
2	Pesos asignados a los descriptores para dos niveles de estrés . .	30
3	Matriz de confusión para dos categorías utilizando RL	30
4	Pesos asignados a los descriptores para tres niveles de estrés .	30
5	Matriz de confusión para tres categorías utilizando RL	30
6	Matriz de confusión para dos categorías utilizando K-NN . . .	32
7	Matriz de confusión para tres categorías utilizando K-NN . . .	32
8	Comparación del porcentaje de aciertos de cada modelo	32
9	Planificación del proyecto	35
10	Desglose de los costes materiales	36
11	Desglose del coste total del proyecto	36
1	Weights of the input parameters for two level classification . .	45
2	Weights of the input parameters for two level classification . .	45
3	Accuracy of the proposed models	45

Abstract

Historically, the term stress has been used to describe the different reactions that the human body presents under pressure. These unconscious responses occur in order to preserve the individual's integrity and could appear in human signals such as heart activity, sweating or pupil dilatation.

The main goal of this Bachelor Thesis is to provide a system that classifies different levels of stress using physiological signals. For this purpose, an experiment that causes stress has been designed. The experiment consists in four incongruent tests of increasing difficulty performed by 24 participants. Their reaction time and correct answers will be used to approach the performance of the participants to the Yerkes-Dodson curves. The more stress the participant is suffering, the worse is the performance.

Two physiological signals have been acquired aiming to classify different stress levels: Electrocardiogram (ECG) and Galvanic Skin Response (GSR). The sensors used to capture these signals were *SHIMMER3 ECG UNIT* and *Q sensor* respectively. Heart activity is extracted from ECG signal and electrodermal activity is extracted from GSR signal. The features that characterize the heart activity are related to the heart rate, while the features extracted from the GSR signal are related to the sweating.

The classification has been developed applying Logistic Regression (LR) and KNN using leave-one-out cross validation (LOOCV). Two classifications will be proposed: relax and test stress levels and relax, concentration and anxiety levels.

After obtaining the results, the stress levels have been successfully classified using a KNN model with $K = 3$. The accuracy of the system is 75.38% for two stress levels and 63.08% for three stress levels, achieving the main goals of the thesis. According to the LR model, it can also be concluded that the features that provide more information about stress are heart rate variability and the area under the curve of the GSR signal.

Keywords: stress, performance, ECG, GSR, physiological, Logistic Regression, KNN, LOOCV.

1 Introduction

Stress is the term used to name the reactions of an individual under pressure. These changes in the organism occur to preserve the integrity of an individual in situations of danger. Some examples of these reactions are sweating, capillar constriction or pupil dilatation. Getting to know how the human body reacts under stressing situations is useful during certain activities that require a quick reaction time.

1.1 Motivation

There is a relation between the performance during a task and the level of stress the individual is suffering [1]. This relation can be seen in situations where the individuals are subjected to high pressure levels. An example can be positions of responsibility, such as pilots or air traffic controllers, where stress can pose a risk to the safety of other people.

On the other hand, high levels of stress can trigger various health problems. Cardiovascular and respiratory diseases, as well as anxiety and depression are common among those suffering from stress.

For these reasons it is necessary to know how the human body reacts to a pressure situation and define how much stress can support an individual before entering a state of anxiety.

1.2 Goal

The **main goal** of this Bachelor Thesis is classifying stress levels. To do so, 24 subjects will participate in a concentration experiments while their sweating and heart activity is being monitorized. The sensors used to acquire data are the *SHIMMER3 ECG UNIT* from *Shimmer Sensing* and *Qsensor* from *Affectiva*.

The **secondary goals** pursued with this thesis are:

- Creating an experiment that encourages participants stress: there has been a literature search on tests used for this purpose and the four most significant experiments have been adapted to be part of the proposed experiment.
- Processing physiological signals: different signal processing algorithms have been implemented to obtain information related to stress.

- Application of machine learning algorithms to classify different stress levels. These algorithms are Logistic Regression (LR) and K-Nearest Neighbors (K-NN).

1.3 Social and economical framework

According to the World Health Organization (WHO) [2], stress is the second cause of heart diseases in developed countries, specially among positions of high responsibility. As the WHO states, stress at workplace is not only a matter of health, but also a risk for the companies. Stress increases absenteeism and reduces the performance of the employees. It also causes accidents due to the lack of concentration in those employees who suffer stress.

1.4 Regulatory framework

As the Ley Orgánica de Protección de Datos de Carácter Personal establishes, any personal data used for research purposes needs the consent of the involved individual. The main principles the Ley Orgánica de Protección de Datos de Carácter Personal establishes are the following:

- The purpose of any study that uses the data must be informed and consented by the participants.
- The data must be updated and deleted when no longer useful.
- Before the data is used by third parties, the individuals must be informed and asked for consent.

1.5 Thesis structure

This Bachelor Thesis has been divided in the following chapters. Chapter 3 will be the problem statement, where the experiment will be proposed and explained, as how the data is acquired during the experiment. These signals require a prior extraction process aiming to obtain the characteristics related to stress. This process is explained in Chapter 4. Once processed, the results are shown and compared in Chapter 5. Finally, in Chapter 6, the results will be analyzed and compared. Conclusions and future work will also be presented in this last chapter.

2 Estado del arte

Históricamente, el término estrés ha sido utilizado para describir las reacciones del cuerpo humano ante ciertos estímulos, especialmente aquellos que someten al individuo a altos niveles de presión. Estos cambios en el organismo se deben a la necesidad de mantener la integridad del cuerpo humano y para ello se desatan una serie de respuestas tanto psicológicas como fisiológicas. Algunos investigadores hacen distinción entre dos tipos de estrés: *eustress* y *distress* [3]. El *eustress* es la respuesta positiva ante un estímulo estresante mientras que el *distress* es la respuesta negativa a ese mismo estímulo. El *distress*, o distrés, puede hacer que un sujeto expuesto a una alta presión sienta ansiedad, angustia o se sienta bloqueado ante una situación estresante. Es importante tener en cuenta que el distrés es algo que no puede evitarse, pero se puede aprender a manejarlo y a prevenirlo.

Dicho estrés se muestra mediante reacciones fisiológicas que están controladas por el sistema nervioso autónomo (SNA). Éste, a su vez, se divide en dos subsistemas: el sistema nervioso simpático (SNS) y el sistema nervioso parasimpático (SNP). El SNS se encarga de preparar el cuerpo humano para actuar frente a estímulos estresantes mientras que el SNP almacena y dosifica la energía de manera que se garantice el correcto funcionamiento del organismo tras ser sometido a este tipo de situaciones.

Ambos subsistemas controlan funciones que se pueden relacionar con el estado emocional [4]. Esto implica que ciertas personas inestables desde el punto de vista emocional son propensas a sufrir estrés y los problemas derivados del mismo, como enfermedades cardiovasculares, ansiedad o depresión [5]. Pero, ¿es posible medir objetivamente el estrés y el estado emocional?

En 1908, los psicólogos Robert M. Yerkes y John Dillingham Dodson modelaron la relación entre el rendimiento y el estrés que sufre un ser humano a la hora de resolver un problema. Este modelo se desarrolló de manera empírica y explica que el rendimiento aumenta con el estrés, pero solo hasta cierto punto. Una vez superado este punto de estrés en el que el rendimiento es máximo, comienza un estado de ansiedad donde el rendimiento vuelve a decrecer [6]. Esta relación entre estrés y rendimiento se muestra en la Figura 1.

Con el fin de caracterizar la respuesta del cuerpo humano ante situaciones estresantes se han encontrado en la literatura distintos experimentos para

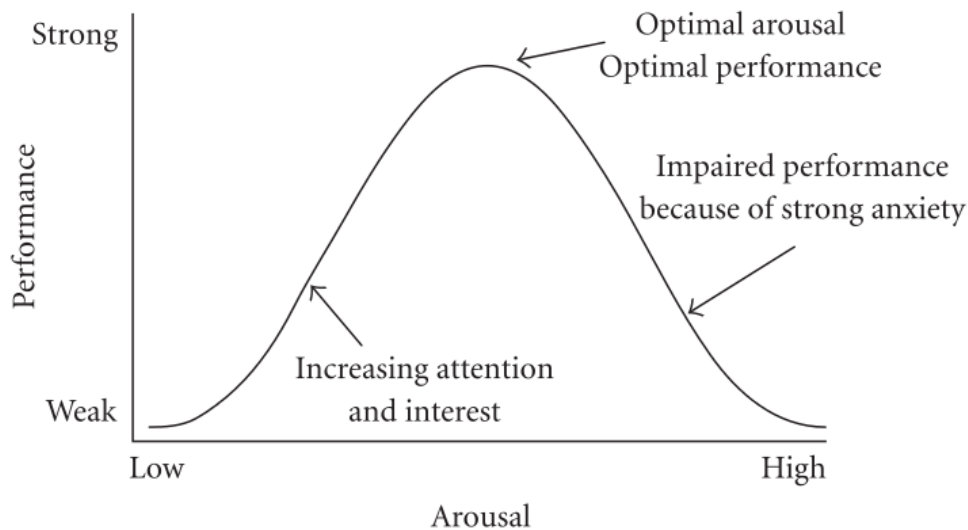


Figura 1: Curva de Yerkes-Dodson [6]

poner a prueba los límites emocionales de un individuo. Esto se consigue capturando las señales fisiológicas del cuerpo humano durante dichos experimentos. Existen para ello sensores que miden estas señales en tiempo real de manera no intrusiva para que los resultados no se vean modificados por la propia medida. Por ejemplo, en caso de que se pretendiera tomar el número de pulsaciones por minuto durante una prueba de habilidad en un ordenador, el pulsómetro no podría ir colocado en la yema de los dedos, ya que el rendimiento de los participantes se vería alterado.

Un ejemplo de los experimentos encontrados es el que propone Jeffrey Brookings [7], donde se sometió a ocho controladores a una simulación de tráfico aéreo. Las actividades medidas fueron el ritmo cardíaco, número de pestaños, movimientos involuntarios de los ojos y el ratio y amplitud de la respiración. Dicha adquisición de datos se realizó con una batería de tests de cinco minutos de duración. La simulación consistía en controlar tráfico aéreo en tres periodos de tiempo, con un descanso de quince minutos entre ambos periodos. El último de estos periodos era un escenario con un espacio aéreo sobrecargado para provocar estrés en los controladores. Además se tomaron medidas de las señales antes de empezar el experimento para obtener los valores de dichas señales en un estado de calma. Los resultados obtenidos concluyeron que cuanto mayor era la dificultad del experimento, menor era el ratio de pestaños por minuto y que el número de respiraciones por

minuto era mayor. Las medidas relacionadas con el ritmo cardíaco, movimientos sacádicos y amplitud de la respiración no mostraban correlación con la dificultad de la prueba.

Aumentando la dificultad del experimento se puede conocer el nivel de estrés máximo al realizar correctamente una tarea [8]. De esta manera se intenta simular la Figura 1, donde el nivel de estrés se puede asemejar al nivel de dificultad. Otro experimento propuesto por Dongrui Wu consiste en tres escenarios de dificultad ascendente conocido como test de Stroop [9]. Teniendo en cuenta el tiempo de reacción y el número de aciertos, los mejores resultados se dieron durante el segundo escenario, donde los niveles de estrés están cerca del máximo de la curva de la Figura 1.

Otro ejemplo puede encontrarse en el campo de la computación. El *Affective Computing* busca adaptar la respuesta de un ordenador acorde al estado emocional del usuario y, para ello, además de medir reacciones fisiológicas, se sirve de señales como expresiones faciales, la voz y los gestos del usuario. Un experimento propuesto por J. Zhai [10] muestra la correlación entre estrés y ciertas señales fisiológicas tomadas cuando los participantes realizaban el test de Stroop. Las respuestas fisiológicas que se tuvieron en cuenta fueron el ritmo cardíaco y la sudoración. Además, utilizando la técnica de Support Vector Machines (SVM), permiten que un equipo diferencie entre estados de calma y de ansiedad de un usuario con un 80 % de acierto.

Por tanto, existen diversas respuestas fisiológicas asociadas al estrés y para no comprometer los resultados de los experimentos se debe tener en cuenta que estas señales se puedan tomar de manera no intrusiva. A continuación se presentan distintas actividades relacionadas con el propósito de este documento.

2.1 Actividad cardíaca

La *actividad cardíaca* se puede medir a través de dos métodos diferentes: *Blood Volume Pulse (BVP)* y *Electrocardiogram (ECG)*.

- **BVP:** es el volumen de sangre que atraviesa los vasos sanguíneos. Se sabe que en situaciones estresantes los vasos capilares se contraen haciendo que el BVP disminuya, como se puede apreciar en la Figura 2. La adquisición de señales BVP se puede hacer mediante fotoplestimografía. El dispositivo emite una luz roja y recibe la que se refleja, que varía con el volumen de sangre que atraviesa los vasos capilares.

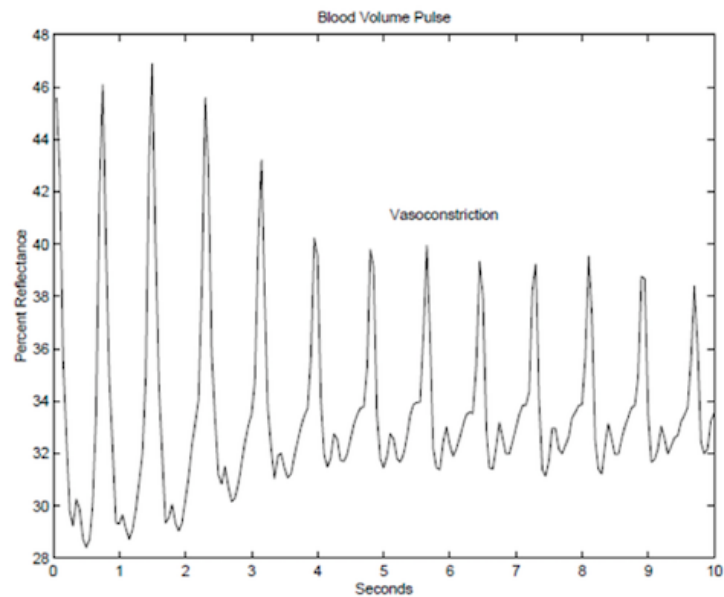


Figura 2: Ejemplo de señal BVP [11]

- **ECG:** es la actividad eléctrica del corazón. El dispositivo utilizado para adquirir esta señal es el electrocardiógrafo, que permite medir la actividad eléctrica de la piel asociada a la actividad cardíaca.

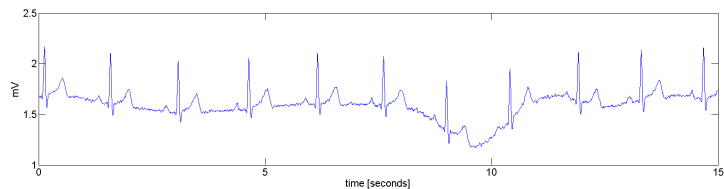


Figura 3: Ejemplo de señal ECG

Como se puede ver en las Figuras 2 y 3, a partir de estas dos señales se puede obtener el Heart Rate (HR), a partir de la distancia entre latidos consecutivos. Ambas señales proporcionan la misma información sobre la actividad cardíaca y, en función de las necesidades, se pueden adaptar a un tipo de sensores u otros.

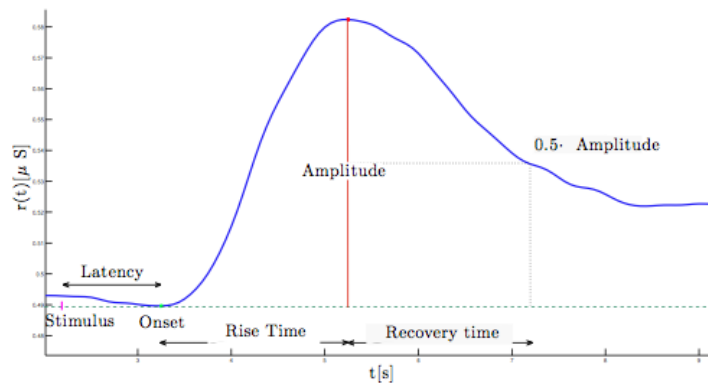


Figura 4: Ejemplo de señal GSR [1]

2.2 Actividad electrodérmica

En situaciones estresantes, las glándulas sudoríparas de la piel segregan sudor, haciendo que las propiedades eléctricas de la piel varíen. Esta señal se conoce como Galvanic Skin Response (GSR). Puede ser capturada mediante dos electrodos que miden la resistencia eléctrica de la piel o su conductancia. La señal GSR tiene dos componentes, la componente tónica y la componente continua:

- Componente tónica: es la componente que varía lentamente y no responde a ningún estímulo.
- Componente continua: aparece como reacción al estrés. Es posible que en ocasiones se pueda dar sin que exista un estímulo.

Un ejemplo de señal GSR puede encontrarse en la Figura 4. Por este motivo se puede obtener la siguiente información de cómo ha reaccionado un individuo frente a un estímulo.

- Estímulo: puede ser un acontecimiento externo al organismo (por ejemplo un cambio de temperatura en una habitación) o interno (por ejemplo un recuerdo).
- Latencia: es el tiempo que transcurre entre el estímulo y la respuesta del organismo. Es importante tener en cuenta que los cambios del organismo se dan en el orden de segundos.
- Respuesta Onset: es el instante en el que la señal llega a su mínimo. Marca el inicio de un cambio en la conductancia eléctrica de la piel.

- Respuesta Pico: es el instante en el que la señal marca su máximo.
- Tiempo de subida: es el intervalo entre la respuesta *onset* y la respuesta pico.
- Amplitud: es la diferencia entre el mínimo y el máximo de la señal GSR. Indica el incremento de la conductancia de la piel debido al estímulo.
- Tiempo de recuperación: es el intervalo de tiempo que tarda la señal en estabilizarse. En este caso se considera que la señal se estabiliza al llegar a la mitad de su amplitud.

La señal GSR posee influencia debido a su forma simple y su facilidad para ser adquirida de manera no intrusiva. No obstante, su principal problema se da cuando dos estímulos ocurren de manera continua en el tiempo. En este caso las respuestas a ambos estímulos se solapan y resulta complicado reconocer el patrón.

2.3 Dilatación de las pupilas

Durante el estado de estrés las pupilas se dilatan y el movimiento de los ojos incrementa su velocidad. Por este motivo es posible conocer el estado emocional de un individuo midiendo estas reacciones fisiológicas.

Una de las maneras de medir el diámetro de las pupilas de manera no intrusiva es *ASL-504 Eye Tracker System*. Este sistema ilumina los ojos del individuo y una videocámara captura imágenes de uno de los ojos. Mediante el reflejo de la luz ultravioleta en la córnea y la pupila, es capaz de calcular el tamaño de la pupila en tiempo real.

Tras un estudio de la literatura de ciertas señales fisiológicas, se ha decidido utilizar las señales ECG y GSR para la realización de este proyecto.

3 Planteamiento del problema

En este capítulo se describe la elaboración de un experimentos en el que se puedan distinguir varios escenarios de estrés con el objetivo de clasificarlos después, así como las herramientas utilizadas a lo largo del trabajo.

3.1 Diseño del experimento

Con el objetivo de incitar estrés, se ha diseñado un test de concentración al que se han sometido 24 participantes de edad media 24.6 años con desviación típica de 3.4 años. Todos ellos se han sometido a la prueba de manera voluntaria. Debido a la Ley Orgánica 15/1999 de Protección de Datos de Carácter Personal no se permite la publicación de la información personal de los participantes, por lo que al inicio de la prueba se les asigna un número identificativo.

El test de concentración al que son sometidos los participantes consiste en cuatro pruebas de atención de dificultad ascendente. La duración de estas pruebas es de cinco minutos, excepto de la última, que dura diez minutos. Todas ellas están separadas por periodos de descanso de cinco minutos. Además, antes del primer test y después del último hay un periodo de descanso adicional de otros cinco minutos. De esta manera, la estructura final de la prueba es la que se muestra en la Figura 5.



Figura 5: Estructura del experimento

Para realizar los distintos test se ha utilizado el software PEBL [12]. PEBL es un plataforma libre que permite crear pruebas cognitivas y de atención y, además, dispone de una batería de experimentos. Esta batería permite modificar distintos parámetros de cada tarea para adaptarlas a las necesidades del experimento.

Gracias a estas características se han podido adaptar cuatro pruebas a las necesidades de este experimento. Estos cuatro test son los siguientes:

3.1.1 Eriksen Flanker Task

Este test es una modificación del *Eriksen Flanker Test* [13]. En esta prueba el participante deberá contestar la dirección que apunta la flecha que aparece en el centro de la imagen. Esta flecha puede aparecer sola o rodeada de otras flechas. Existen tres posibles escenarios que aparecen detallados en la Figura 6.

- **Escenario neutral:** en pantalla tan solo aparece una flecha.
- **Escenario congruente:** en este caso aparecen cinco flechas en pantalla, todas apuntando en la misma dirección.
- **Escenario incongruente:** al igual que en el escenario congruente, en este caso aparecen cinco flechas, pero las que están situadas alrededor de la flecha central apuntan en la dirección opuesta.

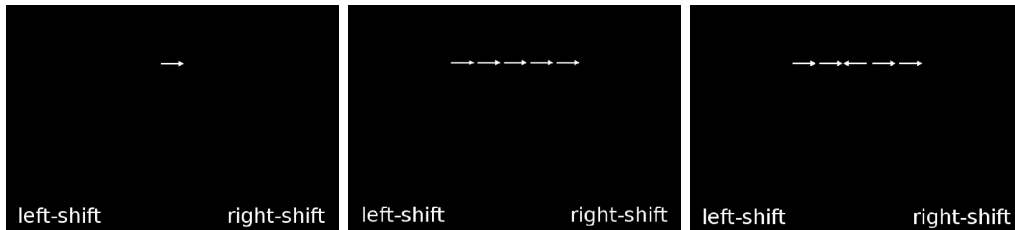


Figura 6: Escenarios del test Eriksen Flanker Task [13]

Para adaptar la prueba a las necesidades del experimento, es decir, hacer que la dificultad se incremente progresivamente, se han modificado la probabilidad de escenarios neutrales e incongruentes, que disminuyen o aumentan respectivamente a lo largo de la prueba. Además, cada veinte intentos el tiempo que tiene el usuario para contestar se reduce desde los 800 milisegundos hasta 250 milisegundos.

3.1.2 Math Processing Task

Esta prueba es una modificación del *Math Processing Task* [14]. Consiste en una serie de ecuaciones en las que el participante debe contestar si el resultado es menor o mayor que cinco. Esta ecuación puede tener entre dos y cinco términos Figura 7.

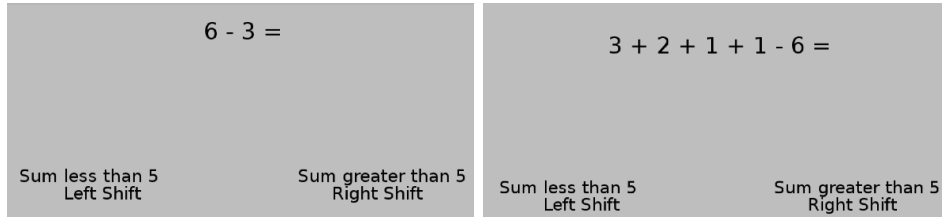


Figura 7: Escenarios del test Math Processing Task

Al igual que en el resto de los test, se han modificado ciertos parámetros para adaptar esta prueba a las necesidades del experimento. Para ello, los primeros escenarios consisten en sumas y restas de no más de tres términos, mientras que en los últimos escenarios las ecuaciones tienen cinco términos. También se reduce el tiempo del que disponen los participantes para contestar, desde 1500 milisegundos hasta 500 milisegundos.

3.1.3 Clock Test

El tercer test al que se someten los participantes es una modificación del *Clock Test* incluido en la batería de experimentos de PEBL [12]. Consiste en un reloj en el que el punto rojo salta a la siguiente posición cada segundo. Un ejemplo de uno de los escenarios de la prueba se puede apreciar en la Figura 8.

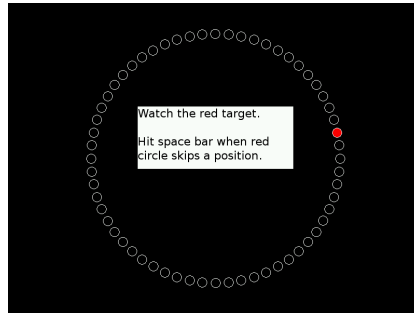


Figura 8: Escenario del Clock Test [12]

En esta prueba los participantes deben pulsar una tecla cada vez que el punto rojo se salte una posición. La probabilidad de que el punto se salte una posición se incrementa a lo largo de la prueba. Además, el radio del reloj decrece cada minuto que pasa, haciendo que las posiciones se solapen como se puede ver en la Figura 9.

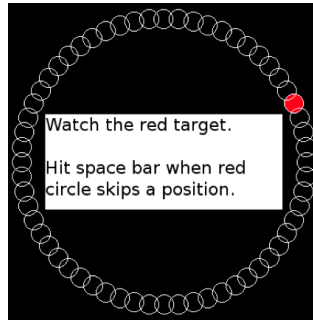


Figura 9: Escenario del Clock Test con solapamiento

3.1.4 Choice Response Time Task

El último test entra dentro de la categoría *Choice Response Time Task (CRT)*, es decir, tareas que requieren una elección en un tiempo de reacción corto [15]. Es una modificación de un experimento de la batería de PEBL [12]. En este caso la prueba consiste en dos partes: estímulo y respuesta, mostrados en la Figura 10.

- **Estímulo:** se muestra una letra al participante durante un breve instante de tiempo. Acto seguido es sustituida por un conjunto de símbolos aleatorios.
- **Respuesta:** se muestran dos letras al participante, una a cada lado del conjunto de símbolos. En esta fase el participante debe decidir cual es la letra que ha visto previamente.

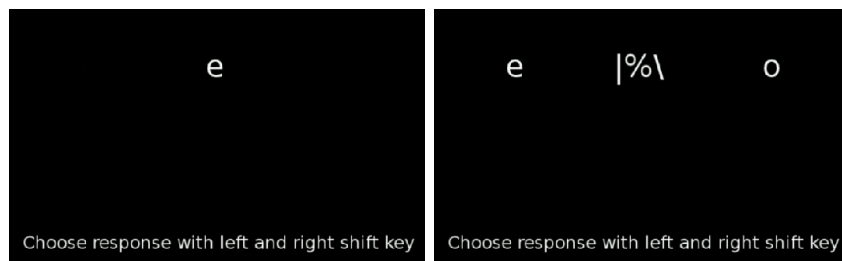


Figura 10: Fases de estímulo y respuesta del Choice Response Time Task

Cabe destacar que esta prueba dura diez minutos frente a los cinco minutos de duración del resto y además hay que tener en cuenta que es la última. El objetivo perseguido con esto es que sea el test más frustrante para incitar ansiedad en los participantes. Para aumentar la dificultad a lo largo de

la prueba se reduce el tiempo de reacción del que dispone el participante, además de que el número de letras que pueden aparecer en pantalla aumenta progresivamente.

Los resultados de cada participante en cada una de las cuatro pruebas queda grabado en un archivo de texto. Estos archivos quedan guardados con el nombre del test y el número identificador del participante. De estos archivos se extrae la información individual acerca del rendimiento del participante. Estos datos son los que se muestran en la Tabla 1.

ID	Test 1	Test 2	Test 3	Test 4
1	Resp. Correctas	Resp. Correctas	Resp. Correctas	Resp. Correctas
	T. Reacción	T. Reacción	T. Reacción	T. Reacción

Tabla 1: Datos extraídos para cada participante

- **Respuestas correctas:** es un vector binario en los que cada valor está asociado a una de las respuestas que ha dado el participante-
- **Tiempo de reacción:** es un vector en el que se muestra el tiempo en milisegundos desde que aparece el estímulo en pantalla hasta que el participante responde.

Teniendo en cuenta que la dificultad de los cuatro tests que componen el experimento es creciente, a partir del tiempo de reacción y del número de respuestas correctas, se intenta asemejar el rendimiento de los participantes con la Curva de Yerkes-Dodson, mostrada en la sección 2 en la Figura 1.

3.2 Adquisición de datos

Es importante tener en cuenta que, para no comprometer la fiabilidad de los experimentos, las señales fisiológicas deben ser tomadas de la manera menos intrusiva posible. Se ha optado por utilizar estas dos señales a lo largo de todo el experimento puesto que existen dispositivos pequeños y precisos capaces de medir ECG y GSR.

3.2.1 Electrocardiogram

Para adquirir la señal del ECG se ha utilizado el sensor *SHIMMER3 ECG UNIT*. Consiste en un módulo principal al que se conectan cuatro electrodos.

Estos electrodos se colocan como se muestra en la Figura 11 y guardan la información capturada en una tarjeta SD. Para extraer los datos del dispositivo, este se coloca sobre una base conectada por USB a un ordenador.

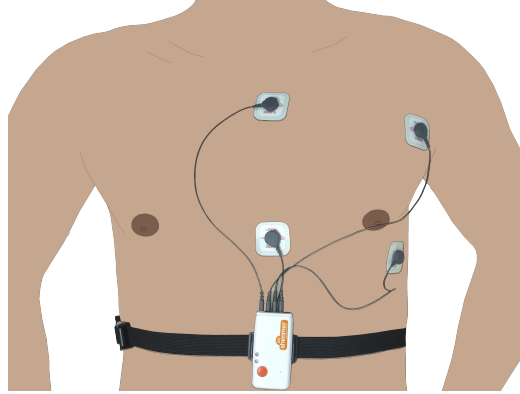


Figura 11: Colocación del sensor *SHIMMER3 ECG*

La extracción de datos del dispositivo se realiza con el software *Consensus* de *Shimmer Sensing*. Esta aplicación organiza los datos en una hoja de cálculo.

3.2.2 Galvanic Skin Response

Para adquirir la señal GSR se ha utilizado el sensor *Q sensor*. Este dispositivo es colocado en la muñeca de los participantes como se muestra en la Figura 12. Dispone de un interruptor que permite marcar eventos que es utilizado para señalar el inicio y fin de las distintas pruebas que conforman el experimento.

La extracción de datos del *Q sensor* se realiza mediante una conexión USB. Es necesario el software *Q* de *Affective*. La grabación de la señal GSR se realiza en un archivo de texto separado por comas.

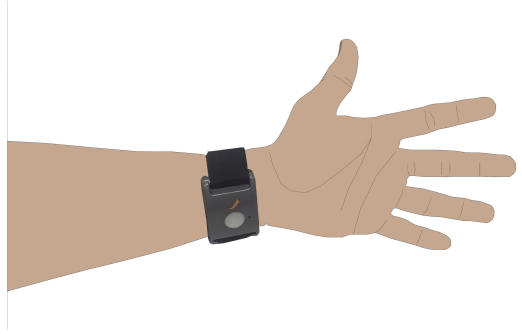


Figura 12: Colocación del sensor Q sensor

4 Procesado de señales

En este capítulo se explican los pasos seguidos para obtener señales de las que poder extraer características. Consiste en una breve explicación del preprocesado de las señales, así como una descripción de las características extraídas de cada una de las señales adquiridas.

4.1 Procesado de los datos

Con el fin de facilitar el procesamiento de la información, las señales han sido sometidas a una manipulación para mejorar su calidad. Este preprocesado se ha realizado con la herramienta **MATLAB**.

Para determinar el nivel de estrés y el rendimiento de cada participante se tienen en cuenta el tiempo de reacción ante cada estímulo y el número de respuestas correctas de cada test. Cuanto mayor sea el tiempo de reacción y más errores cometa, mayor será el nivel de estrés del individuo y menor será su rendimiento.

Debido a que cada prueba tiene sus propios tiempos de reacción máximos y dentro de un mismo test existen distintos niveles de dificultad se ha optado por normalizar todos los tiempos de reacción. Para ello se ha calculado el cociente entre el tiempo de reacción de cada respuesta y el tiempo máximo del que dispone el participante como se muestra en la ecuación 1.

$$t_{norm.} = \frac{t_{Reacción}}{t_{máx}} \quad (1)$$

El número de respuestas correctas se almacena en un vector de valores binarios (0 para los fallos, 1 para los aciertos). Se ha optado por hacer la media en una ventana de cinco respuestas para obtener una gráfica en la que se aprecia mejor el rendimiento a lo largo de una prueba.

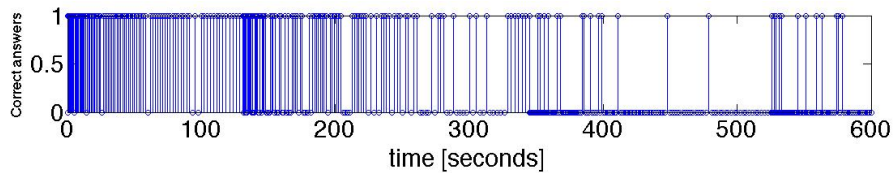


Figura 13: Respuestas correctas e incorrectas a lo largo de una prueba

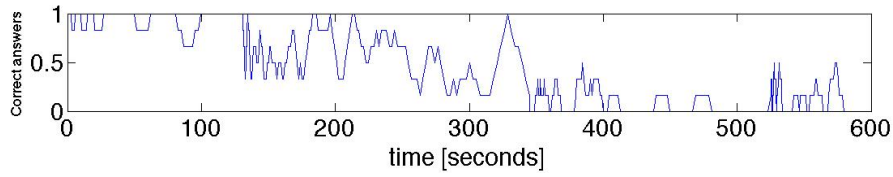


Figura 14: Media de las respuestas en una ventana de cinco muestras.

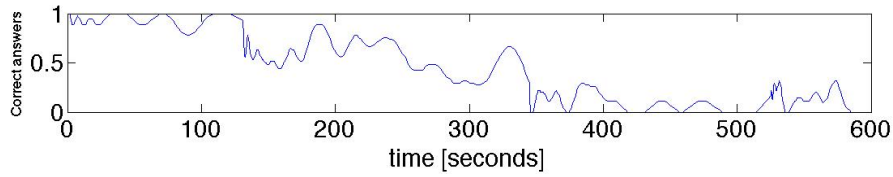


Figura 15: Media de las respuestas suavizadas con un FPB

Una vez obtenida esta señal se le aplica un Filtro Paso Bajo (FPB) para eliminar picos y suavizar la función. De esta manera se evita que picos de baja energía se tengan en cuenta a la hora de realizar la extracción de características.

Por último se calcula el número de respuestas incorrectas a lo largo de la prueba. Esto se obtiene invirtiendo la señal de la Figura 15. De esta manera se puede apreciar gráficamente cómo el número de respuestas incorrectas aumenta según se incrementa la dificultad del test.

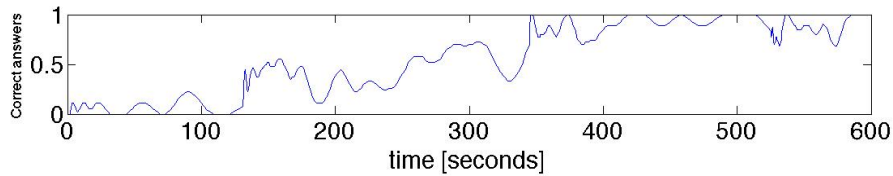


Figura 16: Respuestas incorrectas a lo largo de una prueba

4.1.1 ECG

La señal ECG es común en la vida cotidiana. Al representar la activación del corazón es posible obtener el HR calculando la distancia entre latidos consecutivos o Interbeat Interval (IBI). El algoritmo utilizado [16] detecta los latidos del ECG aplicando un umbral y calcula la distancia los latidos consecutivos haciendo la resta del instante temporal de cada uno. De esta

manera se obtiene el IBI y, a partir de ahí, se puede calcular la frecuencia cardíaca como se muestra en la ecuación 2.

$$HR = \frac{60}{IBI} \quad (2)$$

4.1.2 GSR

La conductancia galvánica de una persona hace referencia a la variación de las propiedades eléctricas de la piel producidas por segregaciones de sudor. Estas incitaciones de las glándulas se ven provocadas por el sistema nervioso simpático (SNS) que controla diferentes funciones del cuerpo como pueden ser ritmo cardíaco, respiración o sudoración.

En la Figura 17 se puede observar un ejemplo de la adquisición de una señal GSR durante 400 segundos. Las franjas marcadas en verde refieren a un etiquetado de respuestas de sistema nervioso ante situaciones de estrés.

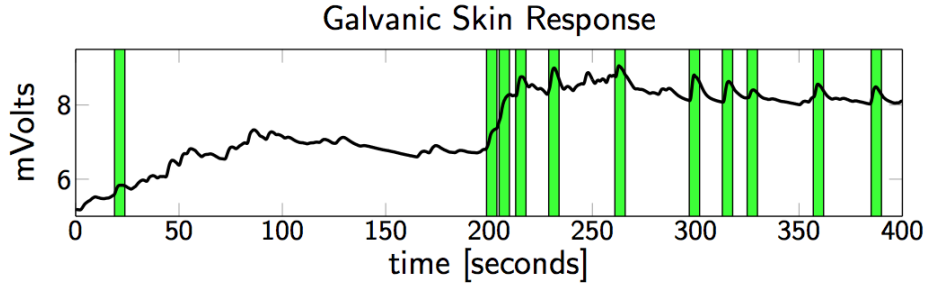


Figura 17: Ejemplo de señal de actividad electrodérmica

En este apartado se explica cual es el procedimiento para detectar reacciones de estrés en las señales adquiridas durante el experimento. Para ello se va a utilizar el método [17] que se encuentra disponible en <http://ledalab.de/>. Dicho modelo supone que la señal se puede dividir de la siguiente manera:

$$s = s_l + r * d \quad (3)$$

Donde s es la señal adquirida, s_l representa la componente continua de la señal, r es la respuesta galvánica comentada anteriormente en la sección 2 y por último d es un tren de impulsos no negativo que refiere a las activaciones de reacciones del sistema simpático nervioso. Dicho señal resultante d es el resultado de una deconvolución de la señal adquirida con una respuesta galvánica fijada. De esta manera, se pueden identificar las respuestas de

sudoración en función del número de impulsos resultantes en la señal d y su amplitud. Además, la parte de la señal continua indica el nivel de atención de cada individuo, siendo mayor la atención para una pendiente positiva y sensación de aburrimiento o dejadez para una pendiente decreciente.

En la Figura 18 se puede apreciar un ejemplo de los pasos seguidos para procesar la señal. El primer paso es la eliminación de artefactos en la señal. Estos artefactos se generan en los instantes en los que los electrodos del sensor no están en contacto con la piel. Para eliminarlos se aplica sobre la señal un filtro de mediana. Este filtro sustituye los valores en un instante por la mediana de los valores que lo rodean en una ventana de 24 muestras.

Una vez eliminados los artefactos se obtiene la componente continua de la señal. Para ello se aproxima la función a un polinomio de primer orden en intervalos de 30 segundos.

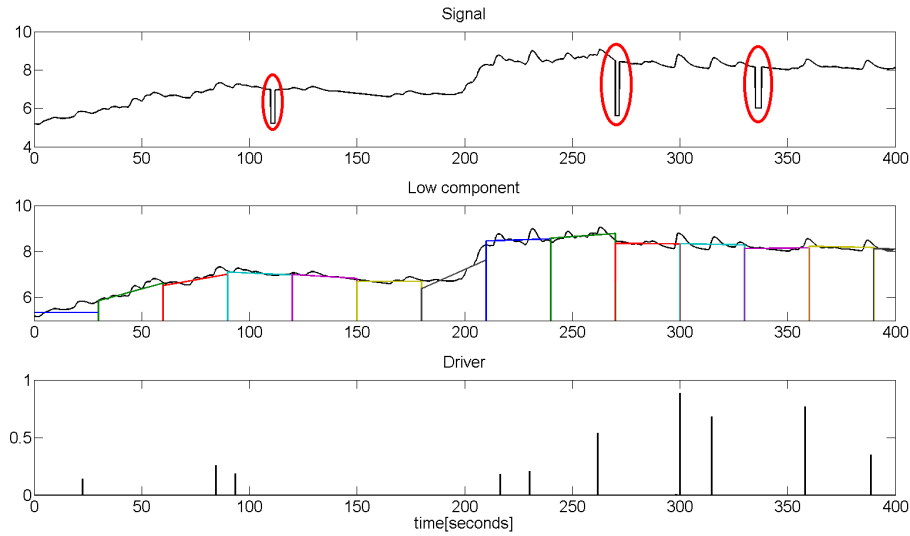


Figura 18: Procesado de señal GSR

Por último, se procede a eliminar la corriente continua y detectar el patrón de sudor en la señal GSR. Para ello se deconvoluciona la señal con una curva modelo como la que se muestra en la Figura 4. Este proceso genera una señal compuesta por deltas de distintas amplitudes que representan cuánto se parece al pulso deconvolucionado.

4.2 Extracción de características

Una vez procesadas las señales se procede a extraer características que permitan clasificar los distintos niveles de estrés de los participantes. Estas características se extraerán de la actividad cardíaca y de la EDA. Las características extraídas de la actividad cardíaca son el HR medio, la variabilidad del HR o HR Variability (HRV), la pendiente del HR. Por otro lado, las características extraídas de la actividad electrodérmica serán el Área Bajo la Curva o Area Under the Curve (AUC) y la pendiente del GSR. Todas estas características se calculan para cada intervalo (test o descanso). Para la extracción de características se ha utilizado la herramienta *MATLAB*.

4.2.1 Actividad cardíaca

- **HR medio:** es el valor medio de todas las muestras dentro de cada intervalo. En la Figura 19 se muestra un ejemplo del HR medio para cada intervalo de test y relajación. La línea roja muestra el valor medio del HR en cada intervalo.

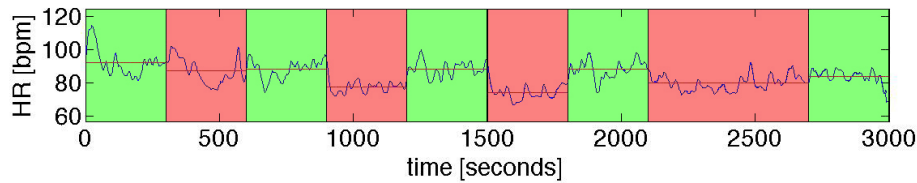


Figura 19: HR y HR medio

- **HRV:** es la varianza de todas las muestras dentro de cada intervalo. Se ha observado que el HRV aumenta en los intervalos de relajación mientras que disminuye en los intervalos de concentración. Un ejemplo se puede observar en la Figura 20.
- **Pendiente del HR:** es el resultado de aproximar la señal HR a un polinomio de primer orden y obtener su pendiente. Esta aproximación se realiza para los intervalos de test y de relajación. En la Figura 21 se puede observar cómo la pendiente, marcada con una línea roja, es positiva en los intervalos en los que la atención aumenta mientras que es negativa en los periodos de descanso.

Las Figuras 19, 20 y 21 representan el HR medio, el HRV y la pendiente del HR, donde los intervalos verdes y rojos representan los periodos de relajación y test respectivamente a lo largo de todo el experimento.

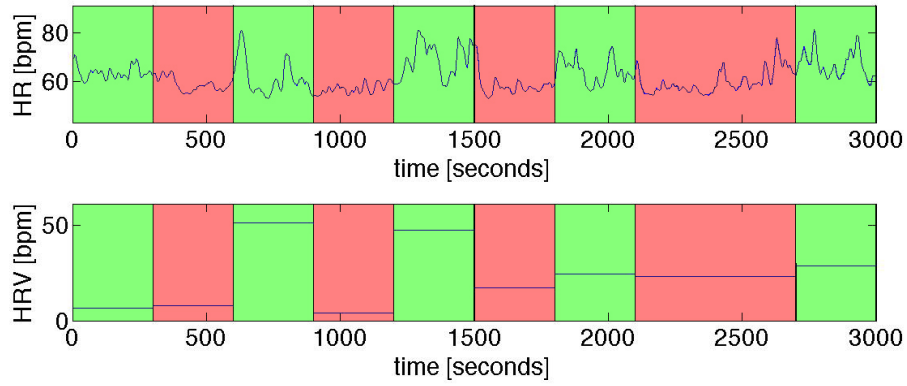


Figura 20: HR y HRV

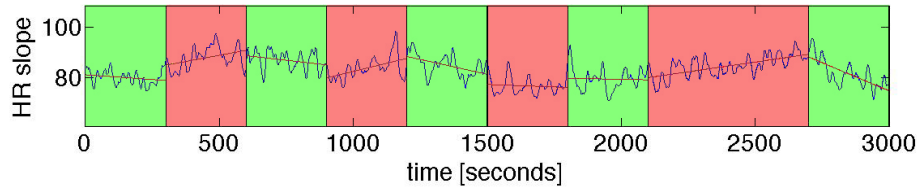


Figura 21: HR y pendiente del HR

4.2.2 Actividad Electrodermica

- **AUC:** coincide con la amplitud de los picos de la señal d . El valor del AUC, por lo tanto, será la suma de las amplitudes de los picos de cada intervalo.
- **Pendiente de la señal GSR:** es el resultado de aproximar la señal GSR a un polinomio de primer orden y obtener su pendiente como se puede observar en la Figura 22. Es la componente continua de la señal y está asociada con el nivel de atención del individuo. Cuanto mayor es la pendiente, mayor es el nivel de atención.

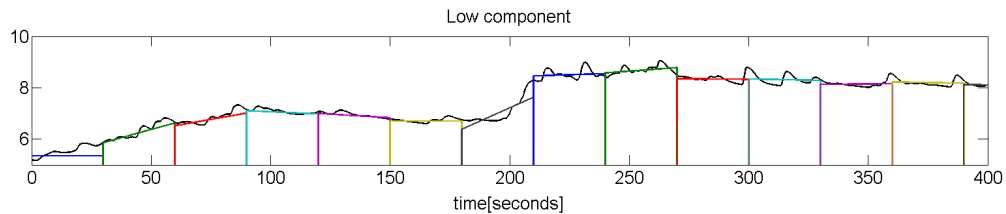


Figura 22: Pendiente del GSR

5 Resultados

En este capítulo se presentan los niveles en los que se clasifica el estrés y los métodos de clasificación utilizados. Además se comprobará la calidad de los métodos propuestos, así como del propio experimento. Para ello se aplicarán dos modelos: regresión logística y K-Nearest Neighbors.

Cabe destacar que, tras realizar las pruebas, se ha detectado que cuatro de los participantes presentaban datos incoherentes. Estos participantes han sido descartados para evitar que los clasificadores obtengan resultados erróneos.

Una vez obtenidos los resultados de los distintos modelos se representarán en su matriz de confusión. Esta matriz muestra el número de muestras de cada clase y la categoría que el algoritmo les ha asignado. De esta manera se obtiene una matriz en la que los valores de la diagonal muestran el número de aciertos mientras que el resto son clasificaciones erróneas.

5.1 Clasificación del estado de ansiedad y rendimiento

Para cada test se van a definir dos intervalos distintos: trabajo y ansiedad. Teniendo en cuenta la Curva de Yerkes-Dodson que aparece en la Figura 1, estos dos intervalos se corresponden con la subida y la bajada de la curva respectivamente. Para encontrar el punto de separación entre trabajo y ansiedad se han tenido en cuenta los resultados individuales de los participantes en el test. Se considera que este instante lo marca el momento en el que los participantes no son capaces de contestar al test de manera coherente.

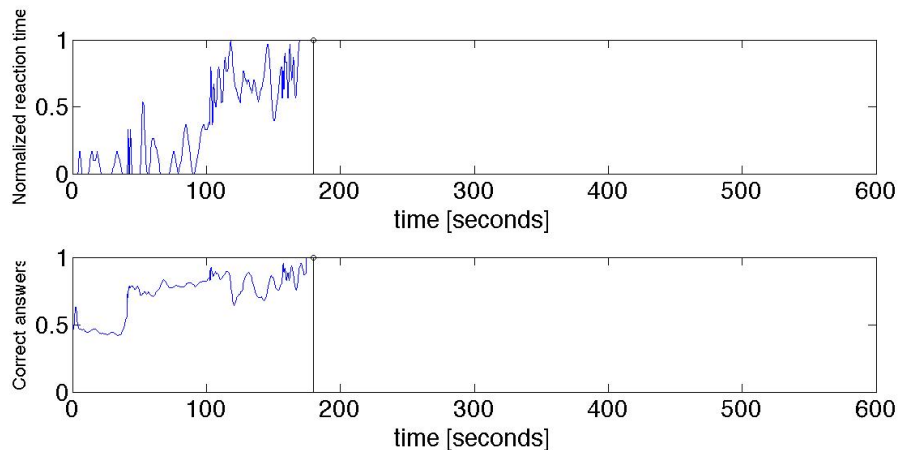


Figura 23: Separación entre niveles de estrés en un test

5.2 Niveles de estrés

El tiempo de corte, por lo tanto, es el instante en el que el participante empieza a sufrir ansiedad. De esta manera se puede hacer una división de tres niveles de estrés: relajación, concentración y ansiedad como se muestra en la Figura 24. El estado de relajación se da durante los descansos, el estado de concentración dura desde el inicio de cada test hasta el instante t_c y el estado de ansiedad dura desde el instante t_c hasta el final de cada test.

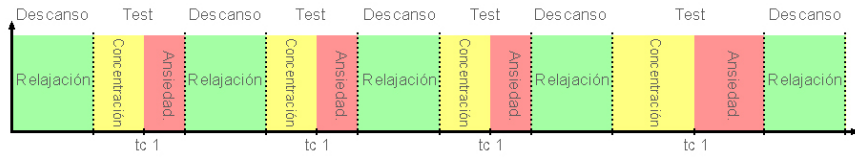


Figura 24: Niveles de estrés a lo largo del experimento

Además de esta clasificación, se ha optado por realizar una segunda clasificación que diferencie entre estado de relajación y de test. El estado de calma es el que transcurre durante los intervalos de descanso mientras que el de test coincide con los intervalos en los que el participante realiza los test.

Por lo tanto, si la clasificación se realiza para tres niveles distintos de estrés se obtienen 5 muestras de relajación, de concentración y 4 de ansiedad por participante. En total se obtienen 100 muestras de relajación, 80 de concentración y 80 de ansiedad.

Por otro lado, al clasificar el nivel de estrés en dos categorías se obtienen 100 muestras totales de relajación y 160 de test, ya que las que anteriormente estaban etiquetadas como ansiedad o concentración ahora son etiquetadas como test.

Una vez delimitados los periodos en los que existen distintos niveles de estrés se procede a etiquetarlos para poder aplicar algoritmos de clasificación supervisados. Esto significa que el propio algoritmo clasifica en distintas categorías en función de lo que ha aprendido en un entrenamiento previo. Para ello extrae descriptores de cada una de las clases etiquetadas. Los métodos usados en este proyecto son Regresión Logística y K-Nearest Neighbors.

5.3 Regresión logística

La Regresión Logística (RL) es una herramienta estadística que permite clasificar una variable en función de las características extraídas. La característica principal de los modelos de RL es la cuantificación de la información que aporta cada una de las características. Los sistemas de RL asignan distintos pesos a cada una de las variables de entrada en función de la información que aporta cada una. De esta manera, aquellas variables con mayor información obtendrán pesos más altos que las que no aporten información o sean redundantes obtendrán pesos bajos.

Para evaluar los resultados del modelo de RL se ha optado por utilizar *cross-validation* o validación cruzada para evaluar los resultados del análisis realizado. La técnica de *cross-validation* consiste dividir en dos subconjuntos los datos obtenidos y utilizar cada uno como conjunto de test o de entrenamiento. Este proceso se realiza de manera iterativa cambiando en cada iteración los datos de los conjuntos de test y de entrenamiento.

Debido a que finalmente se ha trabajado con 20 participantes, el tipo de validación cruzada utilizado ha sido el *Leave-one-out cross validation* (LOOCV). Esta técnica utiliza en cada iteración una sola muestra como conjunto de test y el resto como conjunto de entrenamiento. De esta manera para N participantes se tendrán $N - 1$ muestras en el conjunto de entrenamiento.

La técnica LOOCV devuelve un error para cada iteración. En total se obtienen N valores de error. El error resultante, como se muestra en la ecuación 4, será la media aritmética de los errores obtenidos en cada iteración.

$$E = \frac{1}{N} \sum_{i=1}^N E_i \quad (4)$$

Como se ha explicado en el apartado 5.2, se ha optado por aplicar el modelo RL de dos maneras distintas. En una de ellas se han etiquetado tres niveles de estrés y en la otra se han etiquetado dos.

De esta manera, para clasificar dos niveles de estrés, los pesos que han obtenido las características extraídas se muestran en la Tabla 2.

2 niveles de estrés	Feature	w
ECG	Mean	0.14
	HRV	0.35
	Slope	0.08
GSR	Low c.	0.11
	AUC	0.32

Tabla 2: Pesos asignados a los descriptores para dos niveles de estrés

Una vez aplicado el método RL, la clasificación de las muestras en dos niveles de estrés se muestran en la Tabla 3.

Input	Output class		
	71	29	71,00 %
	44	116	72,50 %
	61,74 %	80,00 %	71,92 %

Tabla 3: Matriz de confusión para dos categorías utilizando RL

En cuanto a la clasificación en tres niveles de estrés, se obtienen los pesos que se muestran en la Tabla 4 para cada una de las características extraídas.

3 niveles de estrés	Feature	w
ECG	Mean	0.09
	HRV	0.37
	Slope	0.03
GSR	Low c.	0.12
	AUC	0.39

Tabla 4: Pesos asignados a los descriptores para tres niveles de estrés

Para esta clasificación, la matriz de confusión obtenida es la que se muestra en la Tabla 5.

Input	Output class			
	60	18	22	60,00 %
	22	41	17	51,25 %
	22	16	42	52,50 %
	57,69 %	54,67 %	51,85 %	55,00 %

Tabla 5: Matriz de confusión para tres categorías utilizando RL

5.4 K-Nearest Neighbors

El método K-Nearest Neighbors (KNN) estima la probabilidad *a posteriori* de que una muestra pertenezca a una clase. Para ello cada muestra es un vector de M dimensiones, siendo M el número de características extraídas. Una vez entrenado el sistema, las muestras de test son clasificadas dentro de la categoría más frecuente entre las K muestras más cercanas. Para calcular la distancia entre la muestra de test y las de entrenamiento se ha usado la distancia euclídea, que se muestra en la ecuación 5.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^M (x_{ri} - x_{rj})^2} \quad (5)$$

Para la elección del parámetro K se ha comprobado el rendimiento para cada valores entre 1 y 10. Tanto a la hora de clasificar en dos categorías como en tres, el valor que optimiza el algoritmo K-NN es $K = 3$ como se puede ver en la Figura 25.

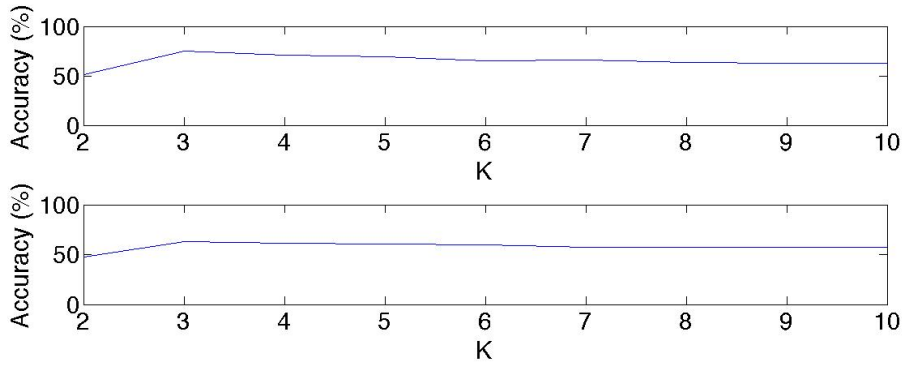


Figura 25: Acierto en función del parámetro K

Una vez obtenido el valor de K con el que se obtienen mejores resultados se realiza la matriz de confusión. De esta manera, la matriz de confusión para dos categorías es la que se muestra en la tabla 6.

Una vez obtenidos los resultados de ambos modelos propuestos se ha realizado una comparativa entre ellos. Para ello, en la Tabla 8 se compararán los resultados de ambos métodos para clasificar dos y tres niveles por separado.

Input	Output class		
	75	25	75,00 %
	39	121	75,63 %
	65,79 %	82,88 %	75,38 %

Tabla 6: Matriz de confusión para dos categorías utilizando K-NN

Input	Output class			
	69	13	18	69,00 %
	20	46	14	57,50 %
	19	12	49	61,25 %
	63,89 %	64,79 %	60,49 %	63,08 %

Tabla 7: Matriz de confusión para tres categorías utilizando K-NN

Niveles de estrés	RL	K-NN
2	71,92 %	75,38 %
3	55,00 %	63,08 %

Tabla 8: Comparación del porcentaje de aciertos de cada modelo

6 Conclusions and future work

In this chapter there will be an analysis about the results obtained in the chapter 5. There will also be a discussion about the experiment and the classification methods that have been used. Finally, future work will be proposed.

6.1 Conclusions

The main goal of this project was to classify different stress levels using physiological signals. In order to achieve this goals two different classification models have been proposed: RL and K-NN. The K-NN model presented an accuracy of 75.38% for two stress levels and 63.08% for three stress levels.

The devices used to acquire the ECG and GSR have not compromised the performance of test subjects. It can be concluded that the devices used are non-intrusive and the results and the signal acquisition were satisfactory, therefore the system is feasible.

According to the results of the RL algorithm, the variables that have more information about stress are the HRV and AUC. The rest of the variables do not contribute with relevant information. After reviewing the results it can be stated that the HRV tends to be constant in stressful situations and more variant during relax periods. It can also be stated that AUC increases during stressful situations. This means that participants tend to sweat when feeling stressed.

In terms of which of the methods proposed has a better accuracy, it can be stated that the K-NN algorithm has a better accuracy. This happens due to the fact that the database of participants only contains 20 test subjects. In order to liken the behavior of the test subjects among each other it can be stated that K-NN is also the most appropriate algorithm.

Overall, the goals proposed in section 1 have been achieved. The system succesfully classifies different stress levels using a K-NN model with LOOCV.

6.2 Future work

Taking this bachellor thesis as a reference, some improvements can be made. In case of continuing with this research, the proposed future work is the following.

- In order to improve dataset it would be very interesting to increase the number of test subjects. This would allow the system to use more training techniques. Designing a new experiment to cause stress on participants would also increase the dataset.
- Currently, all the test samples are treated as they belong to the same test. Aiming to improve the performance of the algorithms, studying how tests cause stress individually is proposed. This would allow the system to determine which tests provide useful information and which ones do not.
- Particularizing the results for each test subject would to obtain the performance-stress curve for each individual. By getting this curves it would be possible to shape de Yerkes-Dodson curve.
- In terms of processing time and optimization, adapting the code to another programming language, such as Python, might improve the performance time.

7 Presupuesto y planificación del trabajo

A lo largo de este capítulo se desarrollará cual ha sido la planificación del trabajo y la duración de cada una de las tareas desarrolladas, así como un resumen del presupuesto necesario para la realización completa del trabajo.

7.1 Planificación

La realización de este proyecto se ha dividido en cinco fases distintas: documentación, diseño, realización del experimento, resultados y redacción de la memoria. La duración de estas fases se muestra en la Tabla 9.

Fase	Tarea	Tiempo (h)
Documentación	Estudio de la literatura	5
	Familiarización con los sensores	5
	Familiarización con las herramientas	6
Diseño	Diseño de experimentos (prototipos)	14
	Realización de pruebas	10
	Desarrollo de scripts	10
Experimento	Realización del experimento	24
	Preprocesado de los resultados	14
Resultados	Procesado de datos	70
	Pruebas de entrenamiento y test	40
	Optimización de algoritmos	10
Memoria	Familiarización con el entorno LaTeX	5
	Redacción de la memoria	72
	Corrección de errores	5
	Presentación	10
TOTAL		300

Tabla 9: Planificación del proyecto

7.2 Presupuesto

El presupuesto necesario para la realización del proyecto se puede dividir en dos partes: costes materiales y costes personales.

Para el cálculo de los costes personales se ha tenido en cuenta que la tabla de honorarios del Colegio Oficial de Ingenieros de Telecomunicación establece unas tarifas de 60€/hora por un ingeniero y por el director del proyecto 150€/hora. Partiendo de estos datos y, teniendo en cuenta que el 15 % de las

300 horas dedicadas al proyecto han sido compartidas con el tutor, el coste del personal es de **24.750€**.

Para el cálculo de los costes materiales se ha tenido en cuenta que la duración del proyecto ha sido de 5 meses y que la amortización del ordenador utilizado, así como de ambos sensores, es de 5 años. Por otro lado, la licencia de MATLAB tiene una amortización de un año. Además de estos gastos, se considera que existe un 20 % adicional de los gastos indirectos. De esta manera los costes materiales quedan desglosados en la Tabla 10 y suponen un total de **323.2€**.

Material	Coste total (€)	Coste proyecto (€)
Ordenador	1500	150
SHIMMER3	448	44.8
Qsensor	400	40
Licencia MATLAB	69	34.5
Gastos indirectos (20 %)		53.8
TOTAL		323.2

Tabla 10: Desglose de los costes materiales

De esta manera, tras aplicar el Impuesto sobre el Valor Añadido (IVA), el coste total del proyecto es de 30.338€ como se puede observar en la Tabla 11.

Concepto	Coste (€)
Personal	24750
Material	323.2
I.V.A. (21 %)	5265.37
TOTAL	30338.57

Tabla 11: Desglose del coste total del proyecto

A Lista de Acrónimos

AUC	Area Under the Curve
BVP	Blood Volume Pulse
CRT	Choice Response Time
ECG	Electrocardiogram
EDA	Electrodermal Activity
FPB	Filtro Paso Bajo
GSR	Galvanic Skin Response
HR	Heart Rate
HRV	Heart Rate Variability
IBI	Interbeat Interval
KNN	K-Nearest Neighbor
LOOCV	Leave-one-out Cross-validation
RL	Regresión Logística
SNA	Sistema Nervioso Autónomo
SNP	Sistema Nervioso Parasimpático
SNS	Sistema Nervioso Simpático
SVM	Support Vector Machines

B Summary: Classification of stress levels applying physiological signals

In this appendix, the bachelor thesis *Classification of stress levels applying physiological signals* will be summarized. It has been written by *Sergio Alcázar Rentero* and supervised by *Antonio Artés Rodríguez* and *Francisco Hernando Gallego*.

B.1 Introduction

Stress is the term used to name the reactions of an individual under pressure. These changes in the organism occur to preserve the integrity of an individual in situations of danger. According to researchers there are two kinds of stress: eustress and distress [3]. Eustress is known as positive stress as it is responsible of the increase of attention. On the other hand, distress is known to cause anxiety when an individual feels overwhelmed.

Stress is controled by the Autonomous Nervous System (ANS). It is in turn divided in two subsystems: the Sympathetic Nervous System (SNS) and the Parasympathetic Nervous System (PNS). The SNS prepares the human body to act in stressful situations while the PNS stores and dosifies energy to guarantee the proper functioning of the system after a stressful situation.

The **main goal** of this bachelor thesis is to classify different stress levels. In order to do so, the **secondary goals** are designing an experiment that causes stress on participants, processing different physiological signals to extract features that help identifying stress and the application of machine learning algorithms to detect stress over time.

The relationship between performance and stress has been found in the literature. The most representative example of it is the Yerkes-Dodson curve 1. It was modeled empirically and represents the performance of an individual during a task of increasing difficulty.

B.2 Experiment design

In order to evoke stress, an experiment has been designed. 24 participants have performed different tests. Their ECG and GSR have been acquired during the performance.

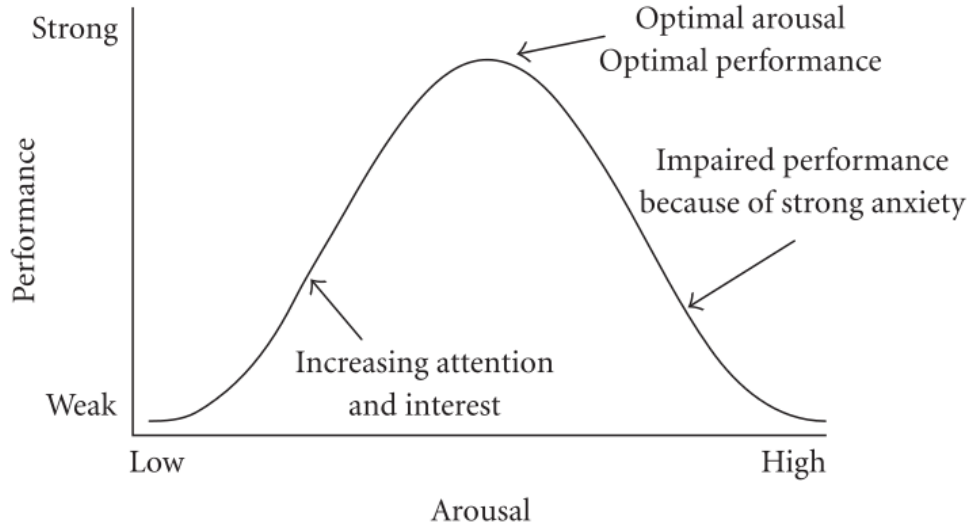


Figure 1: Yerkes-Dodson curve [6]

B.2.1 Experiment

The experiment consists in four tests separated by relax periods as it can be seen in Figure 2. Each of this tests lasts 5 minutes, except for the last one, which is 10 minutes long.



Figure 2: Experiment structure

PEBL [12] is the software selected to create a reliable experiment. It is a free open source software and includes a battery of different tests. The tests that constitute the experiment are adaptations of the battery tests. Thanks to these modifications, an experiment with increasing difficulty can be proposed.

Eriksen Flanker Task

The first test proposed is a modification of The Eriksen Flanker Task [13]. In this test the participant must answer which is the direction the arrow in

the centre of the screen is pointing. As it can be seen in Figure 3, there are three possible scenarios: neutral, congruent and incongruent.

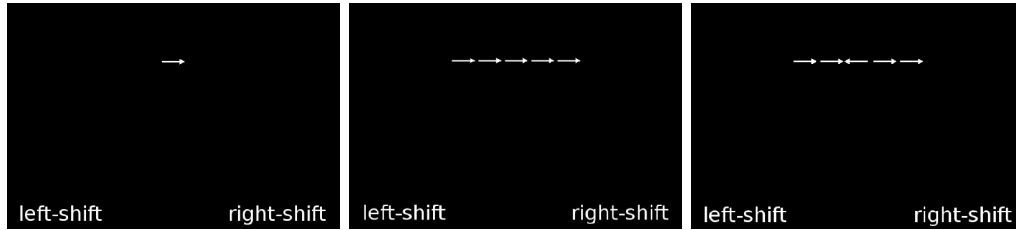


Figure 3: Eriksen Flanker Task Scenarios [13]

Math Processing Task

The second test the participants are required to fulfill is the Math Processing Task [12]. The object of this test is to answer whether the result of the equation that appears on screen is smaller or greater than five. There are four possible scenarios where the equation can have from two to five terms as seen in Figure 4.

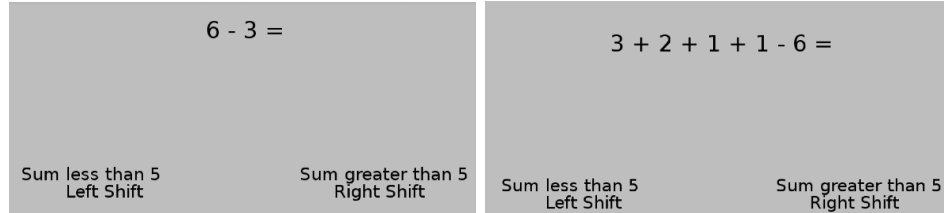


Figure 4: Math Processing Task Scenarios

Clock Test

The third task is called the Clock Test [12]. In this test the individual is asked to press a key if the red dot skips a position. In order to increase the difficulty of the test there are five scenarios where the size of the dots and the radius of the clock change, making it overlay in the hardest scenarios. An example of this can be seen in Figure 5

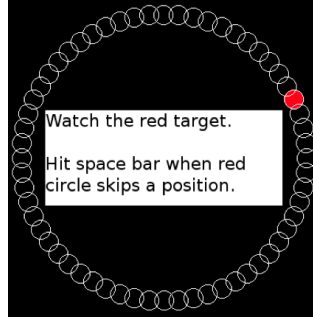


Figure 5: Clock Test scenario

Choice Response Time

The last task is the Choice Response Task (CRT) [15]. In this test the subject is shown a character in the center of the screen for a short period of time. This is called the stimulus phase. After that, two characters appear in the screen and the individual must choose which one was previously shown. This second period is called response phase. In Figure 6.

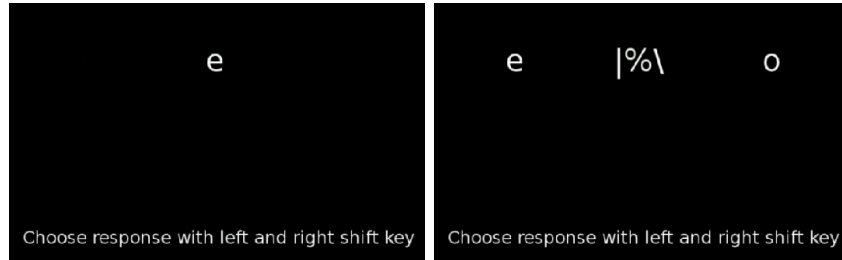


Figure 6: Choice Response Time test phases

B.2.2 Data acquisition

Acquiring physiological signals without altering the performance in the test is key to successfully classify stress levels. The use recording devices that are as less intrusive as possible is necessary to ensure the quality of the experiments. After reviewing different possibilities, using the electrocardiogram (ECG) and galvanic skin response signals (GSR) was decided.

The reasons of utilizing these sensors were their reduced size, making them comfortable for the test subjects, and tracking different activities, such as electrodermal activity (EDA) with the GSR signal and heart activity with the ECG.

The chosen sensors where the *SHIMMER3 ECG UNIT* and the *Q Sensor* to capture the ECG and GSR. In Figure 7 an example of how both sensors are placed can be seen.

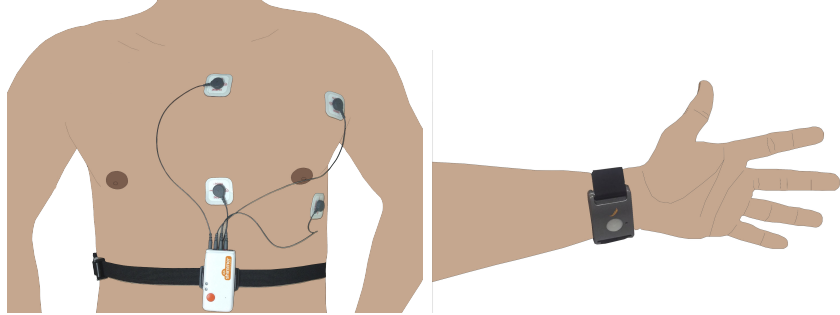


Figure 7: Sensor placement

B.3 Data processing

Once all the tests have been performed, all the data must be processed.

B.3.1 Preprocessing

Before extracting features from the different signals, all of them must be noiseless. This procedure has been divided in two stages: measurement of anxiety and performance and preprocessing of ECG and GSR signals.

Measurement of anxiety and performance and preprocessing

In order to determine the level of stress that participants have and their performance in the different tests, two attributes have been selected. Reaction time and incorrect answers are key to measure how a participant reacts when under stress.

Reaction time is different for each tests and scenario. To study the performance of all the participants in all the tests it has been normalized. This means it has been divided by the maximum reaction time for each scenario.

Incorrect answers also measures the stress level of a participant. PEBL software returns a binary vector of correct answers. In purpose of getting a representative signal of stress, this vector has been applied a low pass filter.

This also removes any outlier, such as having a correct answer in the latest parts of the tests, where correct answers are less likely to happen .

ECG and GSR

Both signals ECG and GSR needed to get preprocessed. By applying a threshold algorithm [16], the interbeat interval (IBI) can be obtained from the **ECG signal**. Obtaining the heart rate (HR) from the IBI can be made with the equation 1. Once the HR is obtained, it has been implemented a low pass filter (LPF) to reduce noise and any other outlier.

$$HR = \frac{60}{IBI} \quad (1)$$

For the **GSR signal** three steps have been followed [17]. First of all, all the artifacts have been removed applying a median filter. Once the signal was clean, the low frequency component has been obtained and extracted from the GSR signal. The last step is deconvoluting the GSR signal and a predetermined model of GSR.

B.3.2 Feature extraction

The features extracted to classify stress can be sorted in two kids: the features of heart activity and the features of the electrodermal activity. All the features are extracted for test and relax periods.

Heart activity

The features extracted from the heart rate are the mean heart rate, the heart rate variability (HRV) and the heart rate slope. It is known that the mean HR and HR slope increase during stressful situations, as well as the HRV decreases.

Electrodermal Activity

For the electrodermal activity the features that have been extracted are the low component and the area under the curve (AUC). It is known that during throughout stressful scenarios the AUC increases. It is also known that the low component of GSR signals increment with concentration.

B.4 Results

Different clasification algorithms have been implemented aiming to classify the different stress levels. This algorithms are Logistic Regression (LR) and K-Nearest Neighbors (K-NN). Both algorithms are supervised methods, therefore all the samples need to be tagged to train the system correctly.

Because of some data being too noisy or missing due to hardware problems, 4 of the 24 participants data have been removed from the experiment.

In order to obtain the best performance from the system, it has been decided that there will be two stress classifications as it can be seen in Figure 8. Relax and test will be one of the cassifications and relax, attention and anxiety will be the second one.

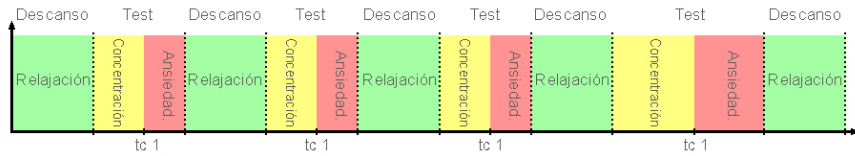


Figure 8: Stress levels

To separate different stress levels within test periods is important to set a cut-off time. This time coincides with the moment when the incorrect answers and reaction time reach the maximum value. This moment represents the maximum of the Yerkes-Dodson curve (Figure 1), which is the point when the performance starts to fall.

Due to the fact that the experiment only counts on 20 participants, the training method used has been a leave-one-out cross-validation (LOOCV). This means each of the proposed methods have been trained 20 times, each one using a test set of just one participant and a train set of 19. To obtain accuracy of the system, the mean of the accuracies for each iteration is computed.

Logistic Regression

The use of Logistic Regression (LR) allows the system to assign a different weight to each of the input parameters, giving greater values to the most

relevant features. This weights are shown in the Tables 1 and 2 for both classifications.

2 stress levels	Feature	w
ECG	Mean	0.14
	HRV	0.35
	Slope	0.08
GSR	Low c.	0.11
	AUC	0.32

Table 1: Weights of the input parameters for two level classification

3 stress levels	Feature	w
ECG	Mean	0.09
	HRV	0.37
	Slope	0.03
GSR	Low c.	0.12
	AUC	0.39

Table 2: Weights of the input parameters for two level classification

K-Nearest Neighbors

The K-Nearest Neighbors (KNN) algorithm computes the probability of a sample belonging to a certain class. In order to do so, it creates a vector of M components, one for each feature, and it is assigned the class of the K closest components. The distance between the vectors is the euclidean distance.

The value for K that has been selected is $K = 3$ as it provides higher accuracy than any other value between 1 and 10.

Result comparison The accuracy for each of the methods proposed and for both classifications are shown in Table 3.

Stres levels	RL ACC	K-NN ACC
2	71,92%	75,38%
3	55,00%	63,08%

Table 3: Accuracy of the proposed models

B.5 Conclusions and future work

The classification of different stress levels has an accuracy of 75.38% for two stress levels and 63.08% for three stress levels. It has been achieved with the KNN system. It can be concluded by the results of the LR model that the most relevant features for stress detection are HRV and the AUC of the GSR signal. Under stress, the HR tends to be constant and the participants are likely to sweat. It can also be concluded that the experiment causes stress on participants, which was one of the secondary goals.

The devices used to acquire the ECG and GSR have not compromised the performance of test subjects. The results and the signal acquisition were satisfactory, therefore the system is feasible.

In terms of which of the methods proposed has a better accuracy, it can be stated that the K-NN algorithm has a better accuracy. This happens due to the fact that the database of participants only contains 20 test subjects. In order to liken the behavior of the test subjects among each other it can be stated that K-NN is also the most appropriate algorithm.

Taking this bachelor thesis as a reference, some improvements can be made. In case of continuing with this research, the proposed future work is the following.

- In order to improve dataset it would be very interesting to increase the number of test subjects. This would allow the system to use more training techniques. Designing a new experiment to cause stress on participants would also increase the dataset.
- Currently, all the test samples are treated as they belong to the same test. Aiming to improve the performance of the algorithms, studying how tests cause stress individually is proposed. This would allow the system to determine which tests provide useful information and which ones do not.
- Particularizing the results for each test subject would to obtain the performance-stress curve for each individual. By getting this curves it would be possible to shape de Yerkes-Dodson curve.
- In terms of processing time and optimization, adapting the code to another programming language, such as Python, might improve the performance.

Referencias

- [1] F. Hernando-Gallego and A. Artés-Rodríguez, “Individual performance calibration using physiological stress signals,” pp. 1–6, 2015.
- [2] W. H. Organization, “Guidelines for the management of conditions specifically related to stress,” 2013.
- [3] M. L. Fevre, G. S. Kolt, and J. Matheny, “Eustress, distress and their interpretation in primary and secondary occupational stress management interventions: which way first?,” *Journal of Managerial Psychology*, vol. 21, no. 6, pp. 547–565, 2006.
- [4] R. W. Picard, “Affective Computing,” *Pattern Recognition*, vol. 20, no. 321, p. 304, 1995.
- [5] P. C. Strike and A. Steptoe, “Depression, stress, and the heart,” 2002.
- [6] D. M. Diamond, A. M. Campbell, C. R. Park, J. Halonen, and P. R. Zoladz, “The temporal dynamics model of emotional memory processing: A synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law,” *Neural Plasticity*, vol. 2007, 2007.
- [7] J. B. Brookings, G. F. Wilson, and C. R. Swain, “Psychophysiological responses to changes in workload during simulated air traffic control,” *Biological Psychology*, vol. 42, no. 95, pp. 361–377, 1996.
- [8] D. Wu, C. G. Courtney, B. J. Lance, S. S. Narayanan, M. E. Dawson, K. S. Oie, and T. D. Parsons, “Optimal arousal identification and classification for affective computing using physiological signals: Virtual reality stroop task,” *IEEE Transactions on Affective Computing*, vol. 1, no. 2, pp. 109–118, 2010.
- [9] J. R. Stroop, “Studies of interference in serial verbal reactions,” *Journal of Experimental Psychology*, vol. 18, no. 6, pp. 643–662, 1935.
- [10] J. Z. J. Zhai, A. Barreto, C. Chin, and C. L. C. Li, “Realization of stress detection using psychophysiological signals for improvement of human-computer interactions,” *IEEE SoutheastCon, 2005.*, pp. 415–420, 2005.
- [11] J. a. Healey, “Wearable and automotive systems for affect recognition from physiology,” p. 158, 2000.

- [12] S. T. Mueller and B. J. Piper, “The Psychology Experiment Building Language (PEBL) and PEBL test battery,” *Journal of Neuroscience Methods*, no. 222, pp. 250–259, 2014.
- [13] C. W. Eriksen and D. W. Schultz, “Information processing in visual search: a continuous flow conception and experimental results,” *Perception & psychophysics*, vol. 25, no. 4, pp. 249–263, 1979.
- [14] W. A. Perez, P. J. Masline, E. G. Ramsey, and K. E. Urban, “Unified Tri-services cognitive performance assesment battery : review and methodology,” 1987.
- [15] R. D. Luce, “Response Times: Their Role in Inferring Elementary Mental Organization,” *Response Times: Their Role in Inferring Elementary Mental Organization*, no. 3, pp. 542–545, 1986.
- [16] J. Pan and W. J. Tompkins, “A real-time QRS detection algorithm,” *IEEE transactions on bio-medical engineering*, vol. 32, no. 3, pp. 230–236, 1985.
- [17] M. Benedek and C. Kaernbach, “A continuous measure of phasic electrodermal activity,” *Journal of Neuroscience Methods*, vol. 190, no. 1, pp. 80–91, 2010.